DOCUMENT RESUME

ED 271 512 TM 860 423

AUTHOR Thorndike, Robert L.
TITLE After 80 Years of G is Testing Going to H?
PUB DATE Apr 86
NOTE 26p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 17-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Aptitude Tests; *Cognitive Measurement; Cognitive Processes; *Intelligence Tests; Learning Theories; Occupational Tests; Predictive Measurement; *Psychometrics; Regression (Statistics); *Test Theory
IDENTIFIERS *General Factor (Intelligence); *Spearmans Two Factor Theory

ABSTRACT

The general ability factor (G), as enunciated by Charles Spearman in the model of cognitive functioning, has been the foundation of psychometric theory and test making practices for 80 years. Through these decades, some psychologists disagreed with this theory, especially Godfrey Thompson and E. L. Thorndike. Nevertheless, various aptitude tests were developed to yield a single score for representing an examinee's general level of cognitive functioning. By 1930 a reaction had set in both on the side of theory and of practice, and tests were developed for specific abilities such as clerical and mechanical skills. In the period between 1940-1980, the concept of general ability was not completely ignored, although it was not emphasized as much; the focus was on multidimensionality of test and specificity in test validity. G refuses to die, however, and has risen like a Phoenix from the ashes. By analyzing several data sets from popular tests, the conclusion can be drawn that the preponderance of the prediction that is possible from any set of tests is attributable to the general ability, and that empirical G lies at the heart of the prediction of real-life performances. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED271512

AFTER 80 YEARS OF G IS TESTING GOING TO H?

Robert L. Thorndike

Teachers College, Columbia University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. L. Thorndike

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860 423

"G"

Robert L. Thorndike

Teachers College, Columbia University

It was just 5 years before I was born that Alfred Binet published the first version of a workable appraisal of human intellect. And at almost the same point in time Charles Spearman enunciated for the first time the model of cognitive functioning as stemming from a pervasive "G," or general ability factor that entered into all cognitive performances.

These were quite independent events, and there is no evidence that Binet and his successors were influenced by Spearman's theorizing, or that Spearman had Binet's test making activities particularly in mind as he developed his formulation of g and the multitudinous s specifics. But the two developments were congruent in that Spearman's theories gave a type of rationale for the melange of cognitive tasks that Binet and his successors assembled in their test batteries. And jointly, the two events provided the impetus for 80 years of psychometric development.

For most of the next 20 years, the focus both of psychometric theory and of test-making practice was on this central general ability. Not all psychologists agreed with Spearman in his concept of a pervasive "mental energy." In particular, Godfrey Thompson at Edinburgh showed that the pattern of test correlations that Spearman relied on as the basis for his concept of "G" could equally well be accounted for by a sampling theory

that stated that different tasks sampled relatively broadly from the host of neural connections that could be made in the cerebral cortex, while in a closely related position E.L. Thorndike viewed level of intelligence as an expression of the number and richness of neural connections.

But whether they explained it as calling upon some pervasive "mental energy" or upon sampling from a supply of neural connections, test makers focussed on instruments designed to yield a single score that would represent the examinee's general level of cognitive functioning. This was true of the original Binet and its many derivatives. It was true of the World War I Army Alpha and Army Beta. It was true of the group tests developed in the decade from 1920 to 1930 by Otis, Terman, Thorndike and others.

But by 1930 a reaction had set in both on the side of theory and of practice. On the theoretical side, Thurstone and his disciples developed the model of cognitive functioning as multi-dimensional, arising from a number of abilities, each of which appeared in only a limited range of tasks. No role was left in their model for a general ability, and research efforts were focussed on defining and expanding the list of so-called "primary mental abilities."

Spearman had come to recognize that the shape of the data made it necessary for him to introduce between his G and the multitude of specifics for particular tasks a certain number of group factors of

intermediate scope, but he thought of these as relatively minor in importance, as compared with the all-pervasive G. Thurstone, on the other hand, ruled G out of the picture, and focussed solely on the several primaries. (We shall see presently that G tended to sneak in through the back door, since Thurstone's primaries turned out not to be independent, but to be correlated.)

More or less concurrent with the shift in theoretical interest to multiple abilities, there was expansion of work to develop tests of more specific abilities. This was presaged by Clark Hull's 1928 book on Aptitude Testing. Present-day psychologists probably only remember Hull as a learning theorist, but before he got involved in learning theory, he wrote what was the first exposition that I had encountered of the rationale and statistics for prediction with multiple and diverse aptitude tests.

Specialized tests for clerical and for mechanical types of jobs began to appear in the 1920's and 1930's. These were often batteries involving speed and accuracy of clerical perception, spatial visualizing, some types of motor coordination, as well as numerical and verbal types of tasks. They fitted into the conception of multiple and relatively specialized abilities as set out by the factor analysts.

The model of ability as multidimensional and relatively specialized achieved its greatest acceptance in the years during and just after World

War II. An outstanding example of a wartime multiple ability battery was the 20-test battery used in the Air Force to select and classify air crew--pilots, navigators and bombardiers. Administered to hundreds of thousands of aircrew candidates and refined through continuous validation studies, it achieved noticeable increments in both direct and differential validity. Since that war, classification batteries for enlisted personnel have been routinely employed in all the Armed Forces.

In the decade after World War II a spate of multiple aptitude batteries made their appearance for use in civilian guidance and employment. Perhaps best known and most widely accepted of these were the Differential Aptitude Battery developed and merchandized by the Psychological Corporation and the General Aptitude Test Battery produced by the U.S. Employment Service. The D.A.T. was prepared to serve primarily as a secondary school guidance tool, while the G.A.T.B. was designed for use in counseling and placement of post-school job applicants.

At the research and theoretical level, the notion of specialized and limited abilities reached its fullest flowering in Guilford's model of the Structure of Intellect, which elaborated a 3-dimensional model of content, operations and products that produced a 4 X 5 X 6 structure containing 120 separate cells. It was felt by Guilford and his disciples that different tests could apropriately be developed for each of these cells, and that this procedure was not only theoretically sound but also likely to be useful in practical affairs.

At the level of application, the impression was widespread among vocational psychologists that each job had its own distinctive pattern of abilities that were required for effective performance , and that the specific abilities required in that job needed to be identified, so that tests could be selected and validated specifically for that job. The impression was fostered by the marked variation in validity coefficients obtained in studies of different jobs, and even in studies of the same job category as reported by different investigators or in different companies. This model of job specificity was cast in concrete by the E.E.O.C. Guidelines, that set forth the requirement that each test must be validated for each job for which it was to be used as a selection device.

In this period from 1940 to 1980, the concept of a general ability was not completely ignored, though it received relatively little emphasis. Several models gave a significant role to a general ability or to a small number of ability dimensions. One might mention Cattell's and Horn's model of fluid and crystalized ability, in which the fluid component represented the biologically given (relatively little influenced by life experiences) and the crystalized represented the overlay of school and out-of-school learning superimposed on the original potential. And relatively close to Spearman's original position is the hierarchical model espoused by Philip Vernon and other British psychologists that pictures an overarching G, with subsidiary factors of verbal: educational, on the one hand, and practical ability, involving mechanical, spatial and perceptual,

on the other. One might also mention, in passing, the notion of simultaneous as opposed to sequential information processing, proposed by Das and given a certain impetus by the Kaufman Assessment Battery for Children. But the "in" thing was multi-dimensionality in test batteries and specificity in test validity.

However, G refuses to die, and has a tendency to rise again like Phoenix from the ashes.

It is important at this point to draw a distinction between G as the purely empirical fact of the almost universally positive correlation between performance on different cognitive tasks and G as a psychological construct. The pervasiveness of positive correlations has been repeatedly shown. Thus, even in the Thurstones' now classic study of their primary mental abilities, 97% of the correlations among the tests in the complete 60-test battery were positive, as were all of the correlations among the 21 tests of a reduced battery that they recommended for possible practical use. A similar result appeared in a World War II research battery administered by Col. Guilford's testing unit to entering Aviation Cadets. In that data set involving 65 variables, only 40 of 2040 correlations were negative. Illustrations could be added ad nauseam. A first centroid factor can be extracted from the correlations among any array of cognitive tests that will usually account for from 40 to 60 percent of the communality of the varied assortment of tests--the shared variance of the tests in the battery. In the Thurstones' study, for example, the percent was 49.6. The

only way that the Thurstones were able to get rid of this evidence of general ability was to allow their so-called primary abilities to become correlated with one another. It is somewhat ironic that the commercially distributed test (the PMA) that the Thurstones provided to serve as a practical measure of these primary abilities has been used by schools largely to yield a single over-all index of cognitive development.

But first factors extracted from different sub-sets of tests do not correspond perfectly. Which, if any, correspond to Spearman's G, sometimes characterized by him as the ability to educe relationships and correlates? And how do we know?

One redeeming fact is that as we increase the number of tests in each battery, the correlation between any two batteries tends to increase. Thus, different batteries tend to converge on more and more nearly the same G. This is illustrated in the Air Force research battery of 65 tests previously referred to. Here, scores on the sum of sub-sets of 8 tests from the battery, in which the average correlation for single tests was about 0.25, correlate to the extent of 0.67 on the average. Sub-sets of 16 show an average intercorrelation of 0.81, while tests 1-32 correlate 0.87 with tests 33-64. The increase is very nearly that called for by the Spearman-Brown Prophecy Formula relating reliability to test length, so it is much as if one were getting an increasingly reliable measure of the same underlying variable. One might argue from this that with almost any sufficiently large and varied set of measures, it is basically the same G

that emerges.

Let us for the present accept the first factor score emerging from a varied battery of cognitive tests as our working definition of G, and ask how much practical utility such a score has. We may phrase the question: How much of the validity that it is possible to squeeze out of a battery of tests, using for each criterion to be predicted a full set of regression weights for that criterion, can be reproduced by a single uniform measure of G? I have analyzed several data sets in order to provide a partial answer to this question.

The point to be remembered is that the true validity of a complex ad hoc set of regression weights can only be determined when those weights are applied to a new and independent cross-validation sample. So the data sets that I have sought have been ones in which a general factor score, which is being viewed as an approximation to G, is compared to regression weights when those weights are applied to a new sample of cases. The model has been one of double cross-validation, i.e., Sample A's weights applied to Sample B, Sample B's weights applied to Sample A, and a uniform set of first factor weights applied to both. I have analyzed three diverse data sets that met my specifications (Thorndike, 1985).

The first consisted of validity and correlational data for the eight Differential Aptitude Tests, as reported in the Manual for that test battery. Validity was an average based on numerous schools and grades

from 8 to 12. The correlations among the 8 tests in the battery were also average values. From the data for boys, regression weights were determined for each of six school subjects. These were then applied to girls, and the weights based on girls cross-validated on boys. In this data set, the first factor G measure predicted 85% as much criterion variance as the cross-validated regression weights--good, but not too good.

A second data set was composed of validity data for 35 different Army Technical Schools. Prediction was from a 10-test enlisted classification battery. Data were available for two separate classes at each school, each with about 250 students, on the average. Thus, regression weights determined on one class could be applied to the other. In this data set, a uniform first factor G score accounted for 90% as much criterion variance as the school-specific cross-validated regression weights. A second factor, orthogonal to the first and appearing to be a clerical versus mechanical contrast accounted for another 6%.

For the third data set, I tried to find one where validities were against job, rather than training, criteria. There had to be validity data for two independent groups in each job, and it was specified that each group must be composed of not less than 50 individuals. The most suitable data set that I found was the voluminous table of validity data contained in the technical manual for the U.S. Employment Service General Aptitude Test Battery. Though validity data were reported there for over 400 different jobs, there appeared to be only 28 that met the criteria that I

had set for inclusion in my analysis.

With the GATB data, the 5 cognitive test scores were used for one analysis, and the 3 motor tests were analyzed separately. Again, the double cross-validation pattern was employed. Focussing first on the cognitive tests, we found that a first factor score, applied uniformly to all jobs, accounted for approximately 120% as much criterion variance, on the average, as the cross-validated regression weights. A similar result was found for the 3 motor tests. A simple sum of the three was superior to the regression weighted combination. With samples of the size used in validation studies of the GATB, the shrinkage in validity on cross-validation was enormous, and developing a tailored set of regression weights for each job seems to be not merely fruitless, but perhaps pernicious.

It is worth noting that the GATB samples were just about the size of those typically found in industrial test validation studies. An extensive survey by Lent, Aurbach and Levin indicated the average sample size to be about 68. That being the case, it is hardly surprising that inconsistent results have been obtained from one study to the next.

When test selection and validation are based on data from large groups accumulated over a period of time, noticeable improvements in validity may be achieved over that possible from a simple first factor estimate of $\underline{G}$. Thus, in the Army Air Force in World War II, an unweighted

composite of the 20 tests in the air crew battery applied to Aviation Cadets accounted for only 65% as much criterion variance as was predicted by the Pilot Stanine developed by extensive validation research. However, it should be noted that Aviation Cadets had been required to pass a screening test that was designed to guarantee general ability comparable to that of an average college sophomore. By contrast, the percent of predictable criterion variance rose to 85 for an experimental group that had not been pre-screened on a general ability measure.

The conclusion to which I am led by these analyses is that the great preponderance of the prediction that is possible from any set of tests is attributable to the general ability that they share. What I have called "empirical G" is not merely an interesting psychometric phenomenon, but lies at the heart of the prediction of real-life performances that is possible from tests.

I would have considered this conclusion rank heresy 25 years ago, when I was still imbued with the spirit of multi-variate prediction stemming from my Air Force experience and the post-war spate of differential test batteries. But in that enthusiasm, I may have lost sight of G. You might say that I am a born-again G-man.

And, of course, I am not alone in this faith. The whole thrust of Hunter and Schmidt's work on validity generalization, over the past 10 years, (1982) is that basically the same ability or abilities are

responsible for the validity of tests over a wide range of jobs. The observed variation in validity coefficients from one study to another is seen as stemming from unstable results in small samples and from differences in the underlying variability of groups and reliability of measures. They would feel quite comfortable, I believe, with the conclusion that the cognitive component of this generalized validity over a wide range of jobs is essentially Spearman's G.

John Hunter (1983) has carried through a wide variety of analyses of the Armed Services Vocational Aptitude Battery. He finds it instructive to analyze the battery at three levels. At the lowest level are the separate tests of the battery. At a second level, these can be grouped into three clusters--designated, respectively, verbal, quantitative and technical. At the highest level, the three cluster scores can be combined into what Hunter speaks of alternatively as "general cognitive ability" or G. Hunter finds that, on the average over jobs, practically all of the prediction of training or job success can be accounted for at the highest and most general level, and that analysis at the level of the clusters or the specific tests can add essentially nothing to that prediction.

Focussing on civilian job categories, and on the tests in the GATB, Hunter (1980) finds a role for two general ability indicies, one cognitive and one motor. Consistent with my more limited analyses, he finds that the specific tests of cognitive skills add nothing to the prediction possible from the general cognitive index, and specific motor tests add

nothing to the prediction from the general motor index. Typically, the motor index had its highest validity for those broad categories of jobs for which the validity of the cognitive index was least, and vice versa.

So far, we have affirmed that psychometric G exists, and that it accounts for most, though perhaps not all, of the prediction of educational and vocational criteria that is possible from a battery of tests of diverse cognitive tasks. The final, and by far the most difficult task is to try to understand the essential nature of G. For this, we must go outside of tests and testing and try to determine the social and especially the biological correlates of individual differences in general cognitive ability. This is an area in which Arthur Jensen has been very active during the past decade, both contributing to and summarizing the body of research. He summarizes inquiry along four distinct lines: (1) the relation of G to speed of elemental information processing, (2) the relation of G to various measures of cortical evoked potential, (3) the relation of G to amount of inbreeding depression, and (4) the relation of G to ethnic group differences.

The number of studies relating speed in processing relatively simple information to measures of general cognitive ability is now sufficiently great so that we can consider the general fact of a relationship to be established. Speed tests themselves tend to be correlated, so that it is possible to extract a "general speed factor." In some of my own work, I have found correlations of about 0.50, on average, between speed in

responding to half a dozen types of stimulus materials. This speed factor appears to be what is correlated with g--and Vernon and Jensen (1984) affirm that the size of the correlation corresponds to the size of the g-loading of specific sub-tests. We have sought to determine whether there is content-specific speed by using two different measures of speed of processing verbal information, two of numerical information, and two of visuo-spatial information, and correlating them with the verbal, quantitative and visuo-spatial level scores of the Cognitive Ability Tests. Approximately 100 children were tested at each of grades 4, 7 and 10. Though the analysis of the results is not entirely complete, they seem to support an interpretation that it is general speed that is related to general cognitive ability, as proposed by Jensen, rather than that speed tasks with a specific type of content are related to level of ability with that same content.

The hope of finding some physio-electrical aspect of cortical response that would relate to psychometric G has been one that has fascinated psychologists and neurologists ever since apparatus has been developed for recording cortical electrical potentials. Ever since 1965, when Barry and Ertl (1965) reported correlations of .88 and .76 between latency scores for the third peak of the evoked potential from a photic stimulus and WAIS full scale I.Q., there has been a flow of research on evoked potential in relation to intelligence. The original Ertl measure failed to be substantiated in a rather larger study of study of school children directed by Frederick Davis (1971), but this has not deterred

investigators from trying out a number of other indices.

Each investigator appears to have his own pet index, and there is a distressing dearth of independent verification of the rather dramatic results that have been reported. I shall mention several indices for which almost implausibly high correlations with G have been reported, but with the caution that these are usually the findings in single studies.

The Hendricksons (1980) claim good correlations for what they speak of as a "string variable"--the total excursion of the electrical response to a stimulus over a specified brief time span. The name derives from the notion that if the record were pulled out into a straight line (as a string could be), the length of that line would represent the total response of the cortex. A long "string" is considered to result from freedom from the damping effect of "errors" in the cortical apparatus, and hence to represent effective cortical functioning.

Schafer (1985) describes two indices--a temporal expectancy index and a habituation index. The first is based on the difference in magnitude of cortical response when presentation of the stimulus comes at random times and when presentation is controlled by the subject. A large decline under the latter condition is reported to be typical of high I.Q. subjects, who are presumably able to make more effective anticipatory adjustments. The second index is based upon the ratio of the response to the early presentations in a series of stimuli to those toward the end of the series.

Greater drop in response was found to be related to higher levels of G, presumably because the more efficient central nervous system shows up in its ability to adjust quickly to unimportant stimulus inputs by ignoring them.

Using a behavioral measure with presumably a neural basis, Brand (1985) offers "inspection time" as a simple index related to G. This is based on tochistoscopic exposure of simple materials followed by a masking stimulus, and is the shortest exposure that enables the subject to report what he has seen with a given proportion of correctness. Correlations of the order of--0.60 are alleged. Thus "speed of intake" is seen as an important component of psychometric G.

In much of his work to explicate the non-test correlates of G Jensen has undertaken to go beyond the general demonstration of a relationship between cognitive ability tests and non-test variables to a more explicit establishment of a relationship between the size of the G loading of the particular item type and the size of the effect. Thus, in a study of the extent of inbreeding depression of ability, Jensen (1983) reports a correlation of about 0.80 over the sub-tests of the WISC between amount of inbreeding depression of score and the G factor loading of that subtest. Similar correlations are reported with speed of information processing measures, with certain evoked potential indices, and with size of black-white score differences. These results, to the extent that they continue to be borne out, indicate that it is predominantly G-ness of a

cognitive task, rather than any of its other features, that determine its non-test correlates. Such findings give G a status above and beyond its statistical origins in the manipulation of a matrix of correlations among test scores, and establish it as having a basic organismic reality.

If one is drawn to the conclusion that G is what makes the predictive wheels go around in most educational and vocational ability testing, one is then led to the question: What is the most effective strategy for measuring G? Two rather sharply contrasting strategies have been followed. The first tries to identify some one testing procedure in which performance is, insofar as possible, purely and solely dependent upon G, uncontaminated by any other influences. The development of Raven's Progressive Matrices exemplified this approach. The test was non-language, and the task modelled closely on Spearman's characterization of G as "eduction of relations and correlates."

The second attack undertakes to sample broadly from a wide range of cognitive tasks, arguing that as the set of tasks is increased in range and number, the importance of factors specific to single tasks or types of tasks will melt away, and the individual's standing on the composite performance will come to depend more and more completely on G. The critical issue is then that the gamut of cognitive tasks be sampled sufficiently broadly and uniformly that no relatively homogeneous cluster of tests introduce significant amounts of variance from factors other than a general cognitive ability. This is the approach that has been

approximated by the Stanford-Binet, the Wechsler series, and many of the group tests, going back to Army Alpha, the Otis Group Test, the Kuhlman-Anderson, and continuing up to the present day. What are the relative merits of the two approaches, and the limitations of each?

A pure test of G would be wonderful, if one could find one. But, alas, this may be as elusive as the pot of gold at the end of the rainbow. Entirely apart from random measurement errors, every type of test task involves some variance specific to it--variance representing familiarity with the type of content, variance arising from specific strategies of attack. As one looks at the correlations among tests, even the most G-loaded has a fair amount of non-chance variance that it does not share with others. No matter how much we may raise the test's reliability by increasing length and more careful item selection--and these are very real practical limits to what we can do in this regard--the specific factor still remains.

What, then, is the strategy for getting the best possible approximation to G? Basically, it is no different than that for predicting any other criterion variable. It is essentially a problem in step-wise regression. One would typically start with the single variable that correlated most highly with G--that is that had the highest G-loading. One would then locate the variable that had the highest partial correlation with G and add it in with appropriate weight. A third and a fourth variable could then be added in turn, until the multiple correlation no longer

increased. The added variables would be ones that not only had a high correlation with G, but also relatively low correlations with other tests already chosen. They would be ones that shared no variance except that which represented this loading with G.

To illustrate how this would work out in practice, I have taken the factor loadings and intercorrelations of the 13 tests of Binet-4 appropriate for a group of 12 to 14-year-olds. The G-loadings are based on a type of confirmatory analysis that assumes a hierarchical structure of G together with group factors. I have also done the same type of analysis with the 9 sub-tests of the Cognitive Abilities Test.

In both of these data sets, the variables that turn out to be the best indicators of G are quantitative tests. These then get supplemented by at least one visuo-spatial test, and then a verbal test. A close approximation to the maximum multiple is reached with 4 or 5 variables. But it is noteworthy that when the full battery is used every subtest receives some positive weight as a predictor of G. Furthermore the equi-weighted complete battery falls only slightly below the regression-weighted composite. I'm not sure whether this is a commentary on the robustness and pervasiveness of G, or on our sagacity (or luck) in selecting the sub-tests to compose the two batteries.

It is my strong belief that the best measure of G--and consequently broadly effective prediction--will stem from a diverse set of cognitive

tasks, chosen to call for relational thinking in a diversity of contexts--what we have tried to provide both in CoGAT and in Binet-4.

Of course, prediction is only one of the uses that one might like to make of ability tests. Another that has received a good deal of attention in the educational literature is adapting instruction to the characteristics of the individual learner. The designation "trait-treatment interaction" was popular a few years back as the basket in which such research was placed.

One might hope that specialized abilities would provide guides for planning the form that instruction should take. Cronbach and Snow (1977) reviewed this literature to see what research had turned up. My impression is that the yield was disappointing, and the one finding that held up in this review was that children low in general cognitive ability required material to be more carefully structured and presented in smaller bites than children higher in G. Snow and Yalow (1982) have taken a more recent look at the field, and doesn't seem to have found much more indication of relationships between specific abilities and specific educational procedures. It is important that we continue to seek, and that test makers provide us the measuring devices that will make such a search possible. In this field also, we urgently need the equivalent of cross-validation, so that we may have confidence in the generalizability of findings. Meta-analyses to extract the kernel of truth from the husks of local circumstances are critically needed. In the meantime, G stands as

one Rock of Gibralter a piece of which will continue to provide a sound investment for our psychometric theory and psychometric practice.

References

Barry, W. and Ertl, J.P. (1965) Brain waves and human intelligence. In Frederick B. Davis (Ed.) Modern Educational Developments: Another Look. New York: Educational Records Bureau.

Brand, C. (1985) Jensen's compromise with componentialism. The Behavior- al and Brain Sciences. 8: 222-223.

Cronbach, L.J. and Snow, R.E. (1977) Aptitudes and Instructional Methods. New York: Irvington Publishers.

Davis, F.B. (1971) The Measurement of Mental Capability through Evoked-Potential Recordings. Greenwich, CT: Educational Records Bureau.

Hendrickson, D.E. & Hendrickson, A.E. (1980) The biological basis of individual differences in intelligence. Personality and Individual Differences. 1: 3-33.

Hunter, J.E. (1980) Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the general aptitude test battery (GATB). Washington, D.C.: U.S. Employment

Service, U.S. Department of Labor.

Hunter, J.E. Validity generalization of the ASVAB: Second report. November, 1982.

Hunter, J.E., Schmidt, F.L. and Jackson, G.B. Meta-analysis: Cumulating Research Findings Across Studies. Beverly Hills, CA: Sage Publications, 1982.

Jensen, A.R. (1983) The effects of inbreeding on mental ability factors. Personality and Individual Differences. 4: 71-87.

Lent, R.H., Aurbach, H.A. and Levin, L.S. Research design and validity assessment. Personnel Psychology, 1971, 24, 247-274.

Schafer, E.W.P. (1985) Neural adaptability: a biological determinant of G factor intelligence. The Behavioral and Brain Sciences. 8: 222-223.

Snow, R.E. and Yalow, E. (1982) Education and Intelligence. In Sternberg, R.J. (ed.) Handbook of Intelligence. New York: Cambridge University Press.

Thorndike, R.L. (1985) The central role of general ability in prediction. Multivariate Behavioral Research, 20, 241-254.

Vernon, P.A. and Jensen, A.R. (1984) Individual and group differences in intelligene and speed of information processing. Personality and Individual Differences. 5: 411-423.